# Use of inverse modeling tools for improved utilization of earth science data in land surface modeling and data assimilation

Sujay V. Kumar, Christa Peters-Lidard, Ken Harrison, Soni Yatheendradas, Joseph Santanello

Hydrological Sciences Branch, NASA GSFC

Earth Science Technology Forum 2010

# Outline

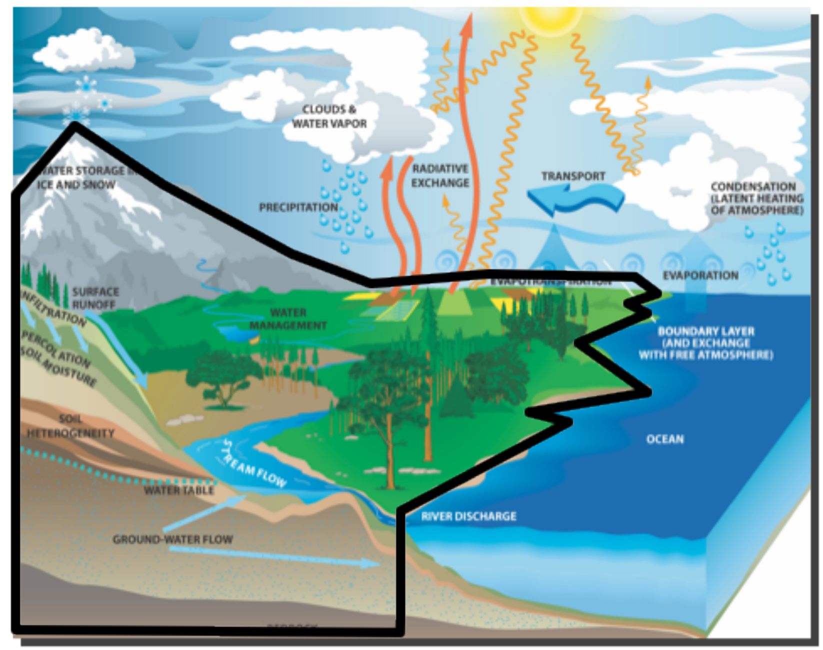
- Motivation
- Land Information System (LIS) background
- Optimization subsystem in LIS – implementation, results
- Uncertainty estimation subsystem in LIS – implementation, results
- Summary

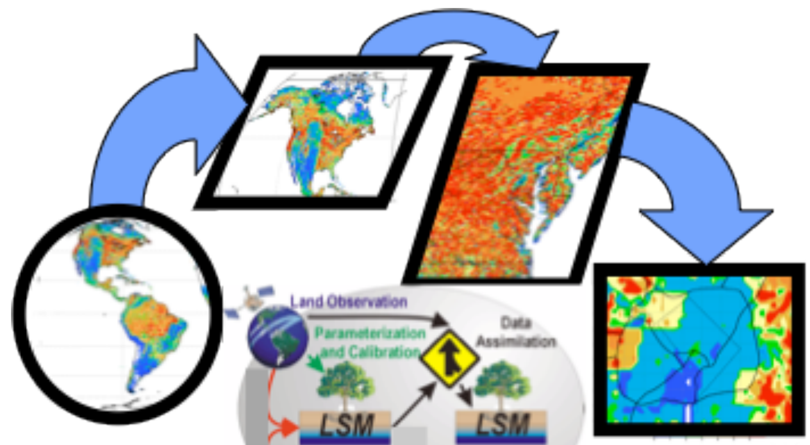
# Motivation

- Develop a suite of inverse modeling tools for improving the exploitation of information content from observational datasets
- Develop a generic optimization infrastructure within National Aeronautics and Space Administration (NASA) LIS (with a suite of optimization algorithms)
  - Improve the land surface model (LSM) forecast accuracy by improving the representation of model parameters
  - Improve the efficiency of data assimilation approaches through unbiased model state prediction
- Develop a suite of uncertainty modeling tools in LIS
  - Quantify the effects of various uncertainties (model parameters, model structural error, measurement errors in input) in the prediction

# Land Information System (LIS)

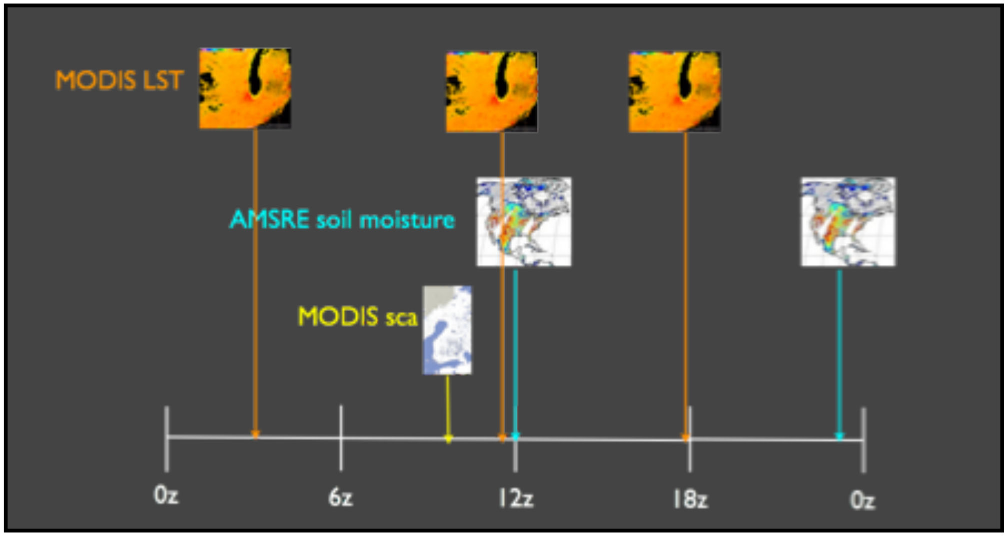
- A system to study land surface processes and land-atmosphere interactions
  - Integrates satellite- and ground-based observational data products with land surface modeling techniques
  - Capable of modeling at different spatial scales
- LIS is used as Problem Solving Environment (PSE) for hydrologic research and as a Decision Support System (DSS) for end use applications
- NASA's 2005 software of the year award
- Used by the Air Force Weather Agency (AFWA) as the operational land surface modeling system



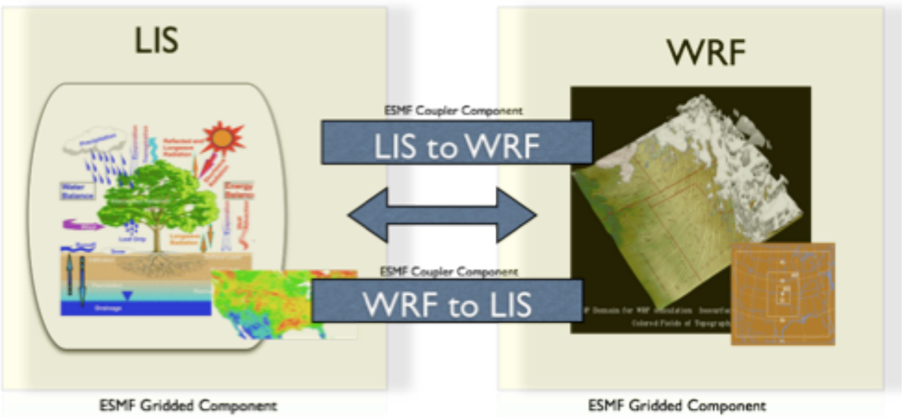



# Key LIS Capabilities

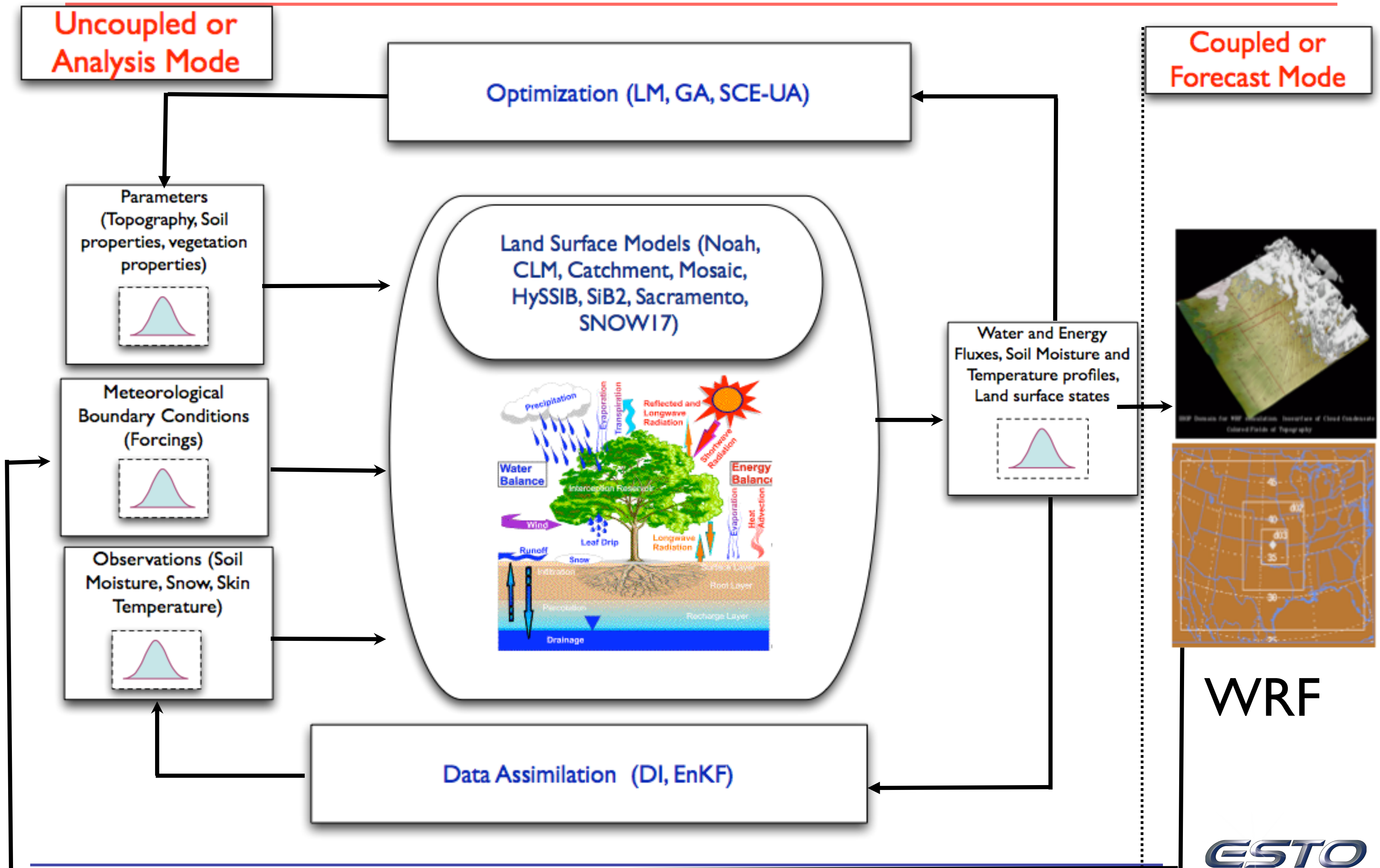
- A comprehensive, sequential data assimilation subsystem based on NASA (Global Modeling and Assimilation Office) GMAO infrastructure
  - Advanced algorithms such as the Ensemble Kalman Filter (EnKF)
  - Interoperable system that allows the integrated use of multiple land surface models, multiple observations and multiple data assimilation algorithms
  - Used in SMAP (Soil Moisture Active Passive Mission) OSSEs and Level-4 assimilation studies



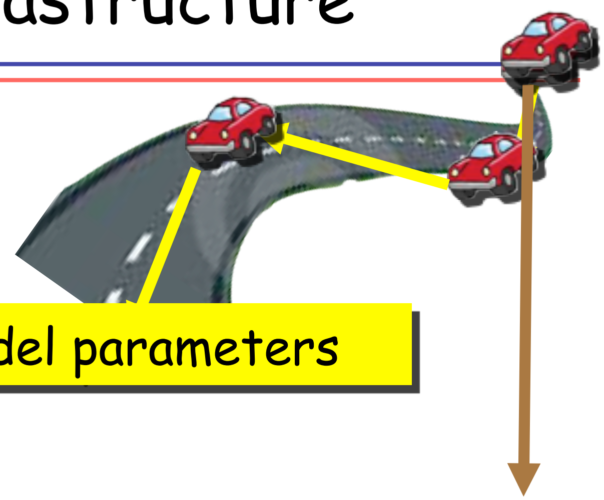

- Coupled land-atmosphere systems that employ LIS as the land surface component (earlier Advanced Information Systems Technology (AIST) funded work)
  - LIS-WRF (Weather Research and Forecasting model)
  - LIS-GCE (Goddard Cumulus Ensemble model)



# LIS modes of operation



# Need for an optimization infrastructure



- Data Assimilation only "adjusts" model states, does not correct inherent model behavior

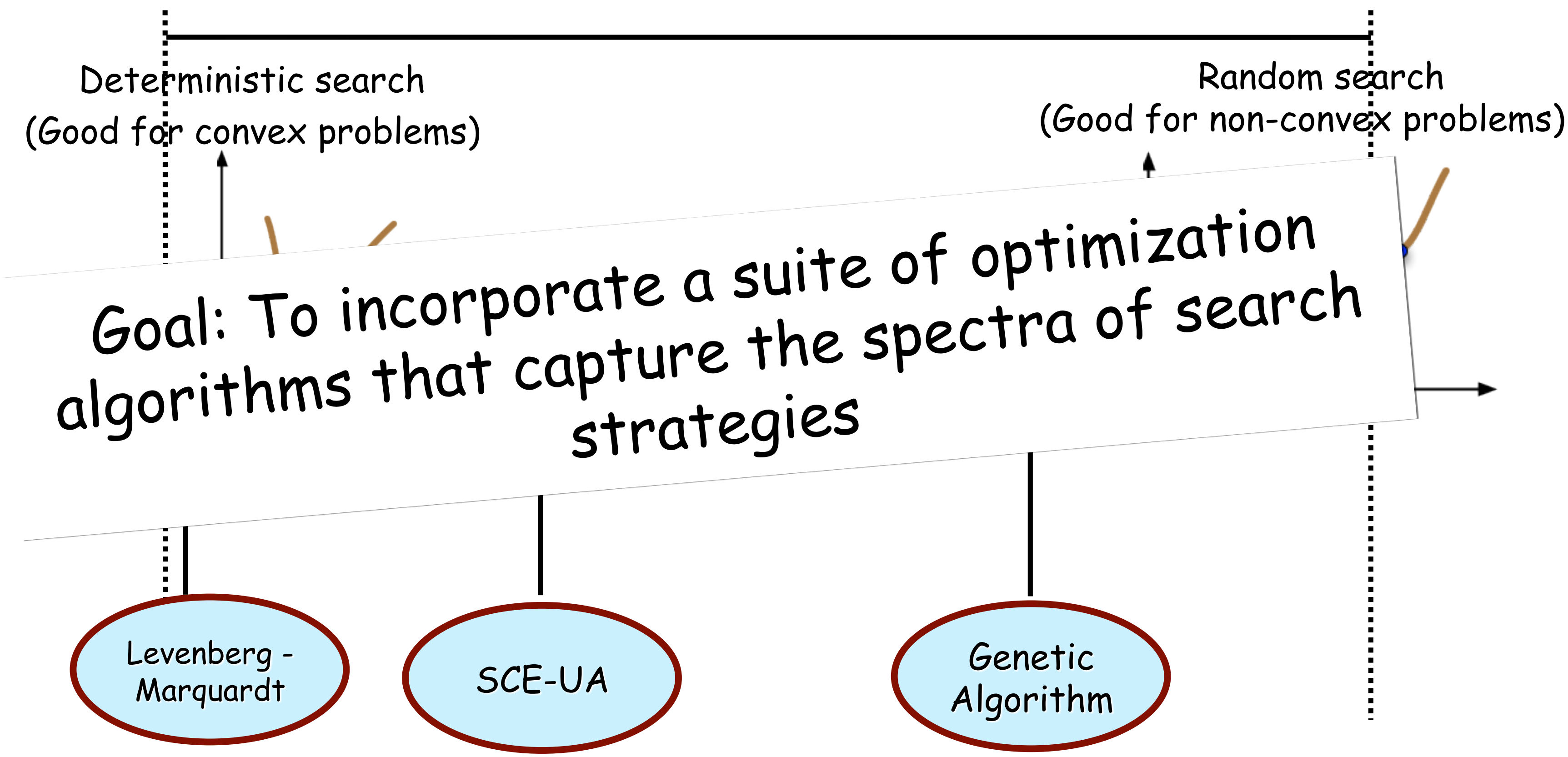
**Use observational information to estimate model parameters**

- Accuracy of model forecasts are sensitive to model parameters
- Some parameters are not easily measurable (hydraulic conductivity, stomatal resistance, aerodynamic resistance) and their relationship at different spatial scales is different
- Existing parameter representation in land surface models are based on tabular results from point samples, and are indirectly represented through categorical data such as soil texture and vegetation class

In addition ...

- Data Assimilation methods rely on unbiased model predictions
- Optimization will enable the adaptive specification of data assimilation error parameters

# Spectra of optimization algorithms/ search strategies



# Parameter estimation

$M$

$\theta$

$\theta = (\theta_1, \theta_2, ...., \theta_N)'$

Parameters (Topography, Soil properties, vegetation properties)

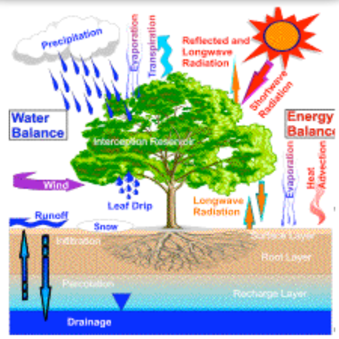
$U_t$

Meteorological Boundary Conditions (Forcings)

$X_t$

States (Soil Moisture, Snow, Skin Temperature)

Land Surface Models (Noah, CLM, Catchment, Mosaic, HySSIB, SiB2, Sacramento, SNOW17)




Water and Energy Fluxes, Soil Moisture and Temperature profiles, Land surface states

$M(X_t, U_t, \theta)$

Observations (Fluxes, soil moisture, snow, skin temperature)

$Y_t$

Model - Actual = Residual $\quad f_t(\theta) = Y_t - M(X_t, U_t, \theta)$

Vector of residuals: $\quad F(\theta) = (f_1(\theta), f_2(\theta), .....f_T(\theta))'$

Measure Z for comparing quality of fit: $\quad Z(F(\theta))$

E.g., “least squares” measure: $\quad Z(F(\theta)) = \sum_{t=1}^{T} [f_t(\theta)]^2$

## Optimization Formulation

$$\min_{\theta} Z(F(\theta))$$

Resulting solution is best fit

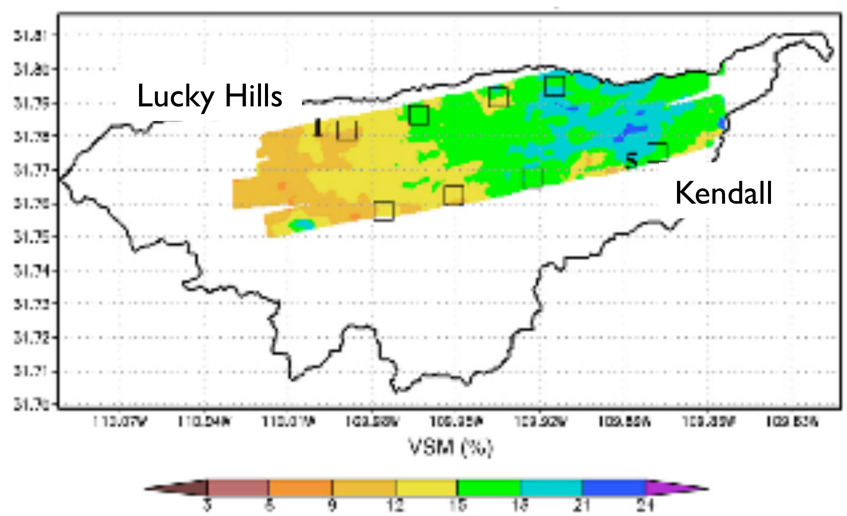
$$\theta^*$$



ESTO Earth Science Technology Office

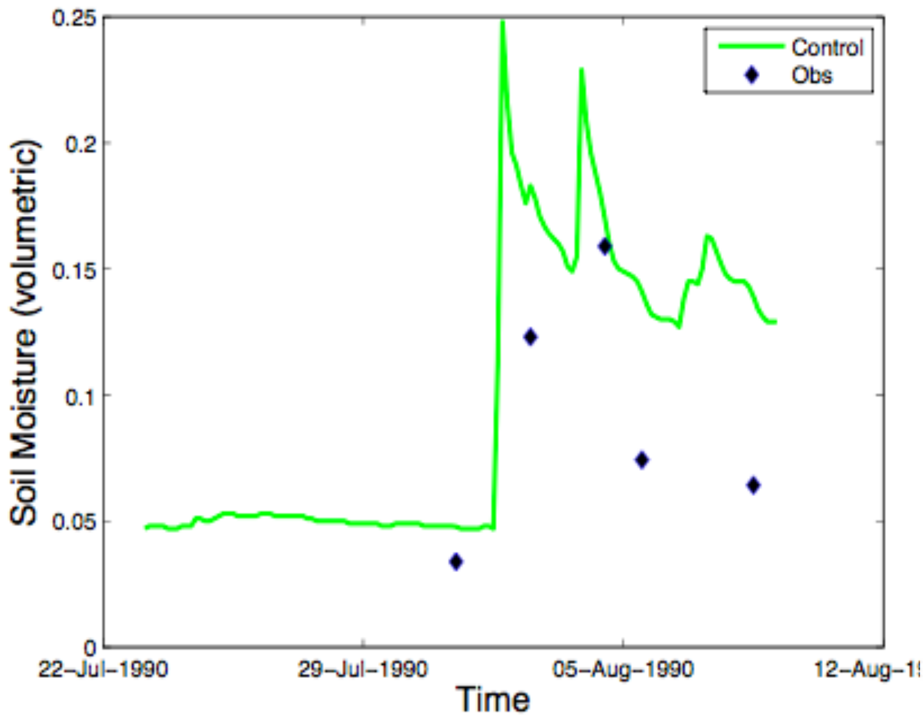
# Test case

- The study used in Santanello et al. (2007) is chosen as the case study to exercise the LIS optimization subsystem.
- Objective: Parameterize soil properties for estimation of soil moisture.
- Observations: Estimates of near surface soil moisture derived from passive (L-band) microwave remote sensing (using NASA's push broom microwave radiometer - PBMR) during six dates during Monsoon '90 experiment (23 July - 9 Aug, 1990) in Southeastern Arizona.
- Noah land surface model employed at 40m spatial resolution across the Walnut Gulch experimental watershed.
- Optimization simulations adjust the sand, clay soil fractions, which in turn control the hydraulic properties through the use of pedotransfer functions (PTF).

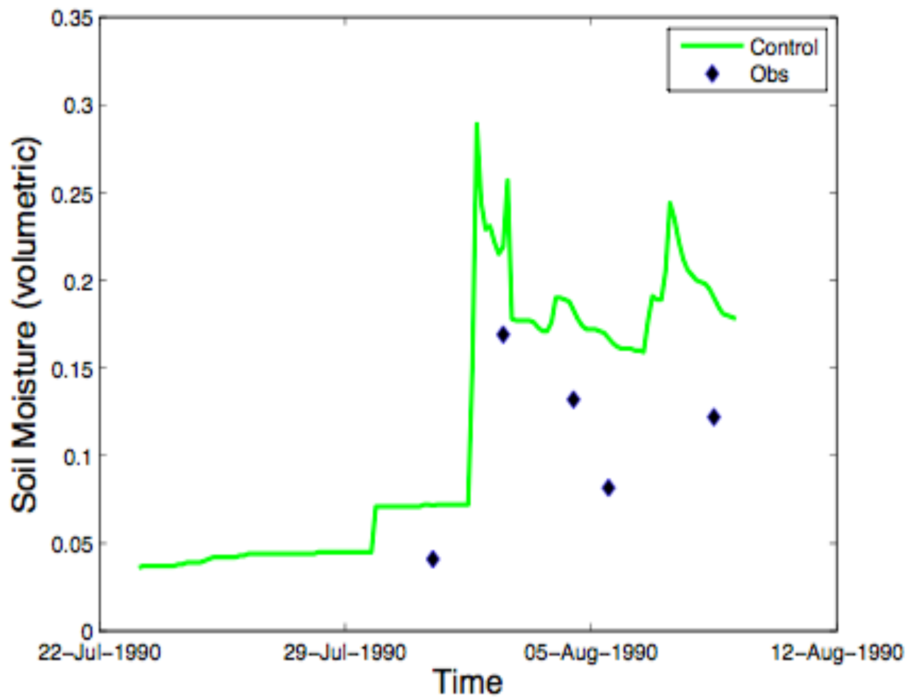


# Model simulations with default parameters

Site 1: Lucky Hills

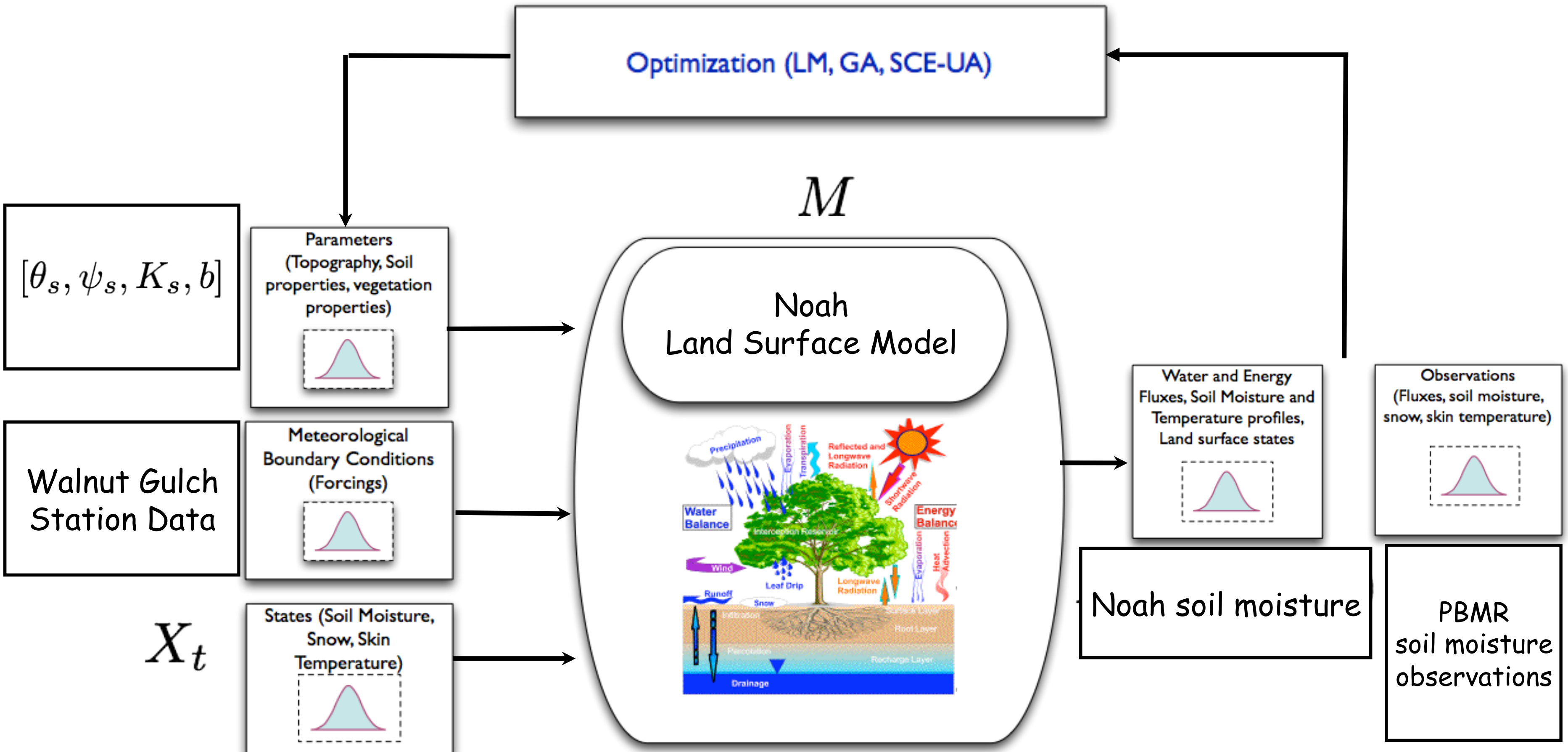



Site 5:
Kendall



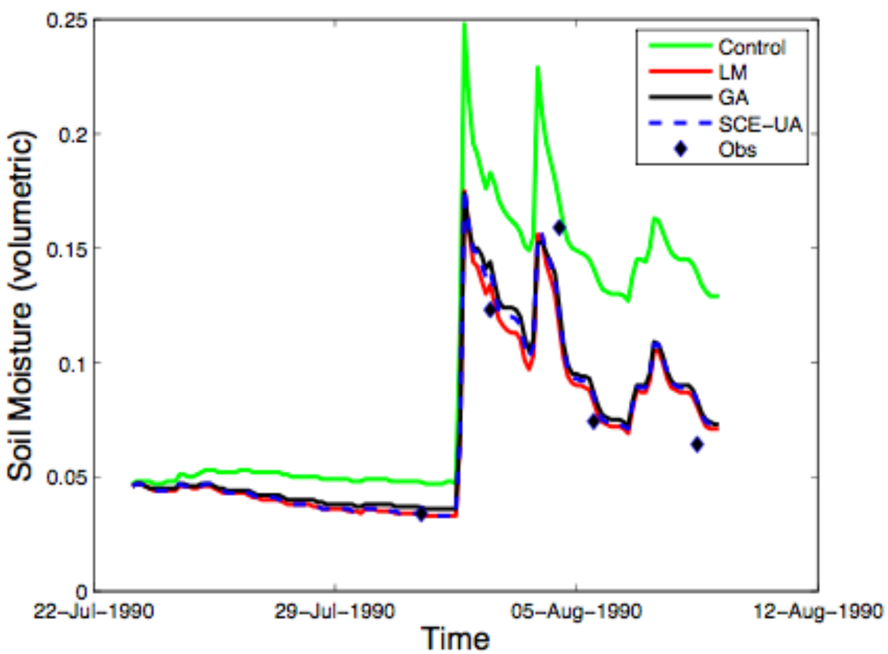
# Parameter estimation schematic for theTest case

$\theta_s$ : porosity $K_s$ : saturated hydraulic conductivity
$\psi_s$ : saturated matric potential $b$ :pore size distribution index

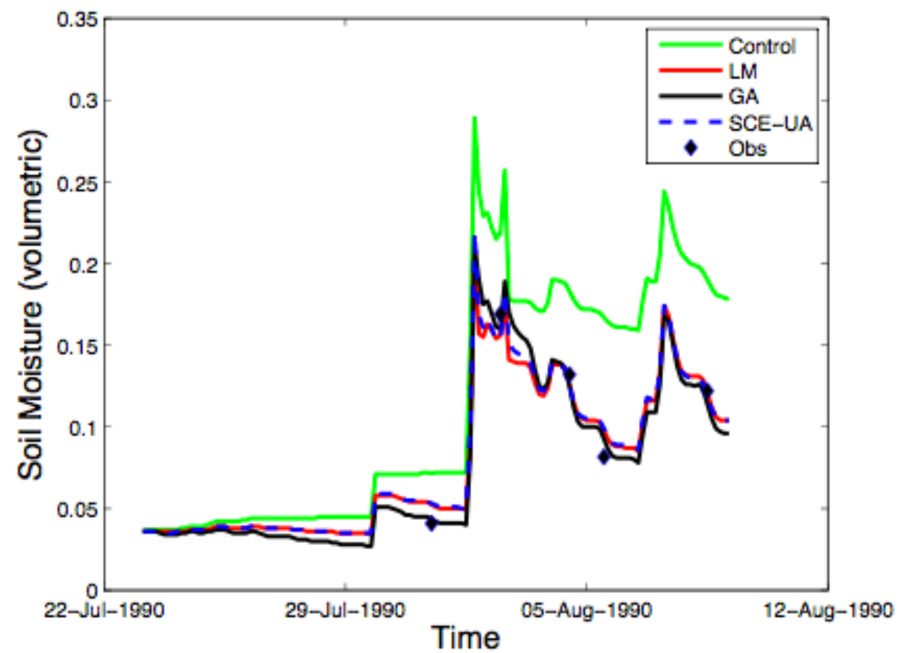


# Comparison of solutions

Site 1: Lucky Hills



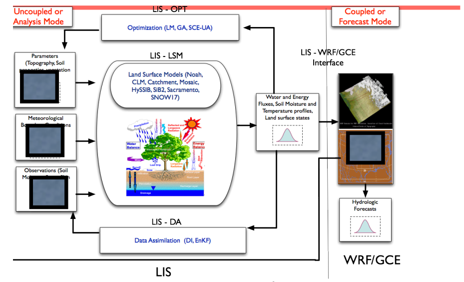

Site 5: Kendall




Earth Science Technology Office

# Need for an uncertainty estimation infrastructure

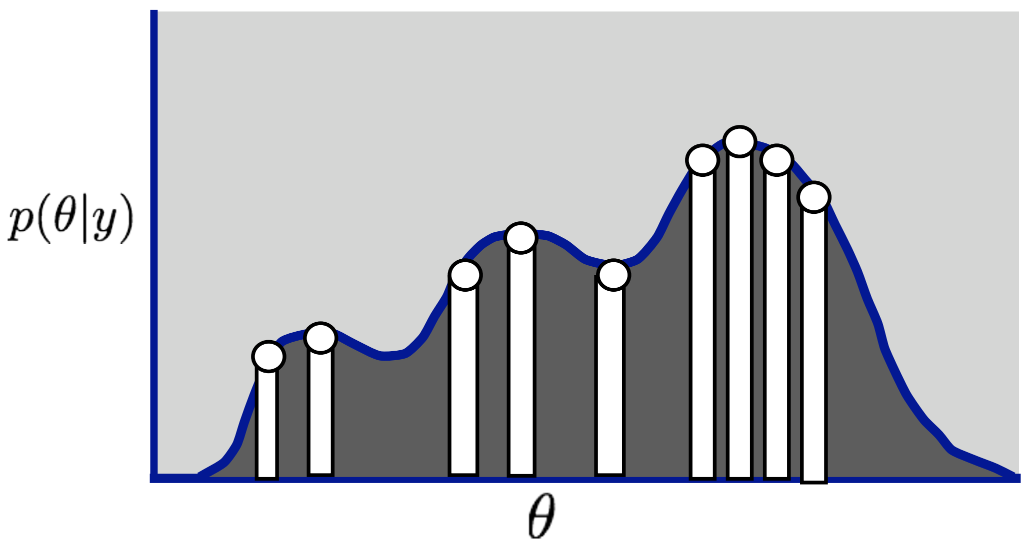
- Land surface model predictions are subject to uncertainties in model parameters, input forcing and model structure - the typical deterministic approach to modeling does not address these issues.




- Beyond the best fit (Equifinality) - Optimization algorithms aim at generating a single optimum (best fit) solution. In contrast, uncertainty estimation tools generate "ensemble" of parameter sets.
  - There may be many different parameter sets within a chosen model structure that may be acceptable in reproducing observed behavior.
- Posterior prediction - Uncertainty estimation algorithms incorporate different sources of uncertainty into probabilistic predictions.
  - Knowledge of uncertainty can help in the risk assessment for decision making (e.g. uncertainty in soil moisture predictions can be used in deciding irrigation practices).
- Preposterior analysis - Bayesian analysis used to investigate the value of data from proposed data efforts. This can assess the value of proposed observing systems in an Observation System Simulation Experiment (OSSE) setting.
  - e.g. What is the added benefit of additional soil moisture observations from SMAP?

# Uncertainty modeling

- Optimization algorithm solves for the best fit.
- In conducting uncertainty analysis, we acknowledge that other solutions have probability.
- How do we correctly generate an ensemble of such solutions (An ensemble that reflects the unknown posterior distribution)?




Bayesian Analysis

# Bayes' rule

- Bayesian inference involves using observations to update/infer the probability that hypothesis (set of parameters) is true.

Posterior probability of $\theta_i$ — Likelihood — Prior probability of $\theta_i$ — Probability of $y$

$$\underbrace{p(\theta_i|y)}_{\text{Posterior probability of } \theta_i} = \frac{\overbrace{p(y|\theta_i)}^{\text{Likelihood}}\overbrace{p(\theta_i)}^{\text{Prior probability of } \theta_i}}{\underbrace{p(y)}_{\text{Probability of } y}}$$

| | |
|---|---|
| $\theta$ | Array of unobservable, uncertain model parameters |
| $\theta_i$ | A particular model "fit" of $\theta$ |
| $y$ | Observations |

$$p(\theta_i|y) = \frac{p(y|\theta_i)p(\theta_i)}{\int_\theta p(y|\theta)p(\theta)}$$

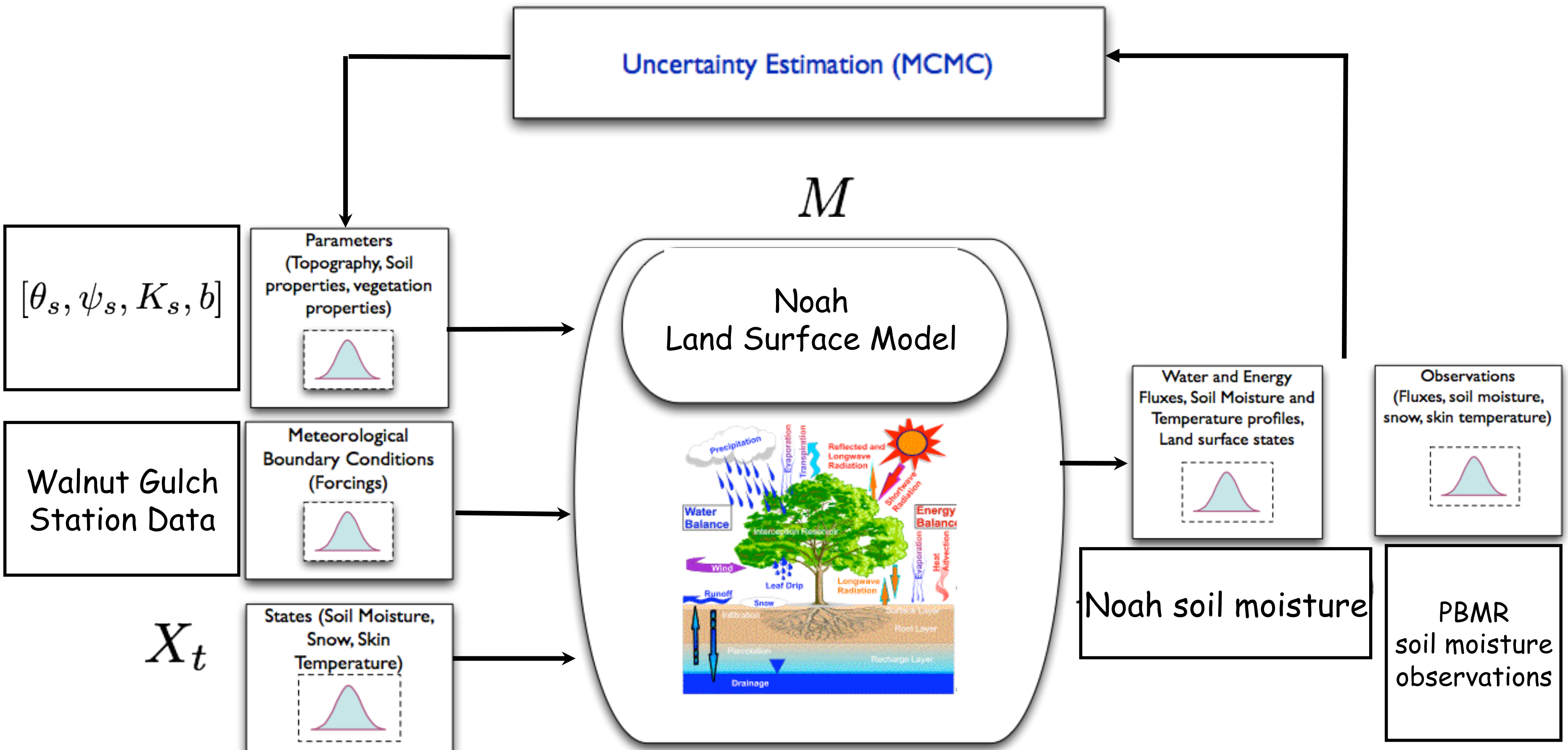
Exceptionally computationally expensive to evaluate with standard integration methods

# Uncertainty estimation algorithms - MCMC

- Markov Chain Monte Carlo (MCMC) method is a stochastic simulation that is based on constructing a Markov chain that has the desired distribution as its target (posterior).
- After a large number of iterations, the frequency of the states of the chain is an estimate of the target distribution.
- The main challenge with MCMC is in reducing the number of iterations required to converge to an equilibrium distribution.
- Different implementations of the algorithm based on different proposal strategies: Metropolis-Hastings, Gibbs, DREAM, etc.

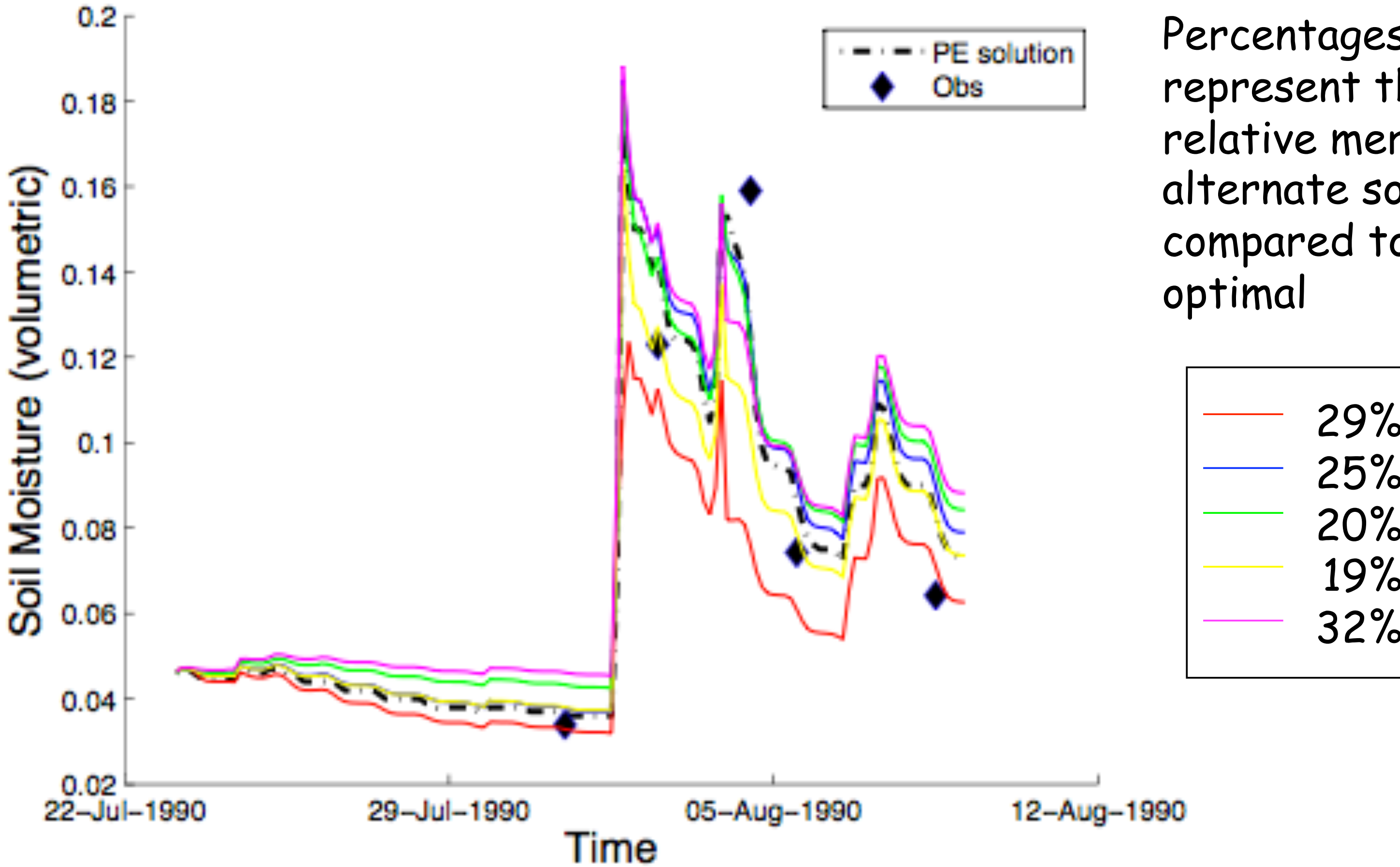
# Testcase for Uncertainty Estimation simulations

$\theta_s$ : porosity $K_s$ : saturated hydraulic conductivity
$\psi_s$ : saturated matric potential $b$ :pore size distribution index



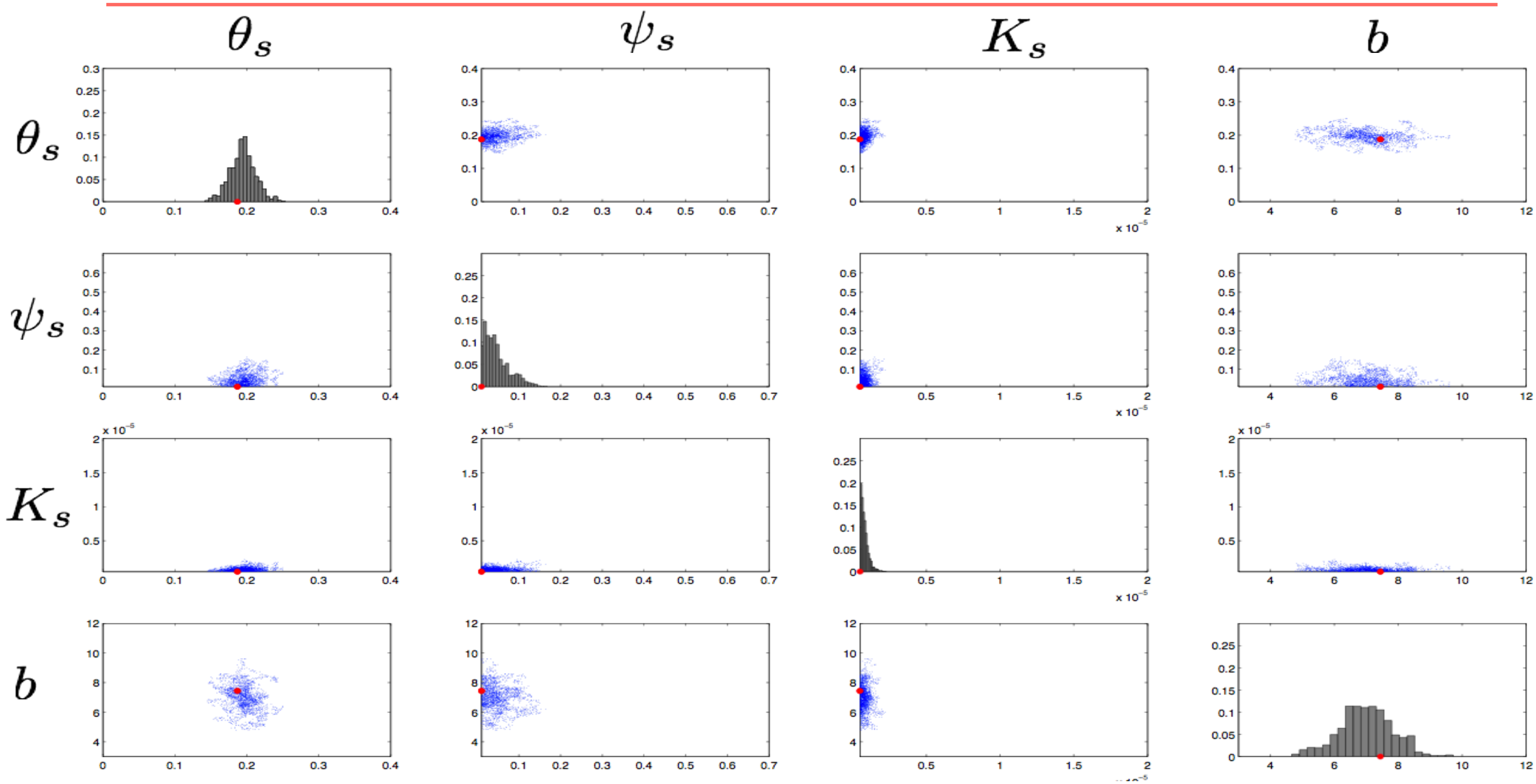
# Beyond the best fit/"Equifinality" of solutions

Sample soil moisture simulations generated by MCMC
(High probability solutions contrasted with the best fit solution)




Percentages represent the relative merit of alternate solutions compared to the optimal

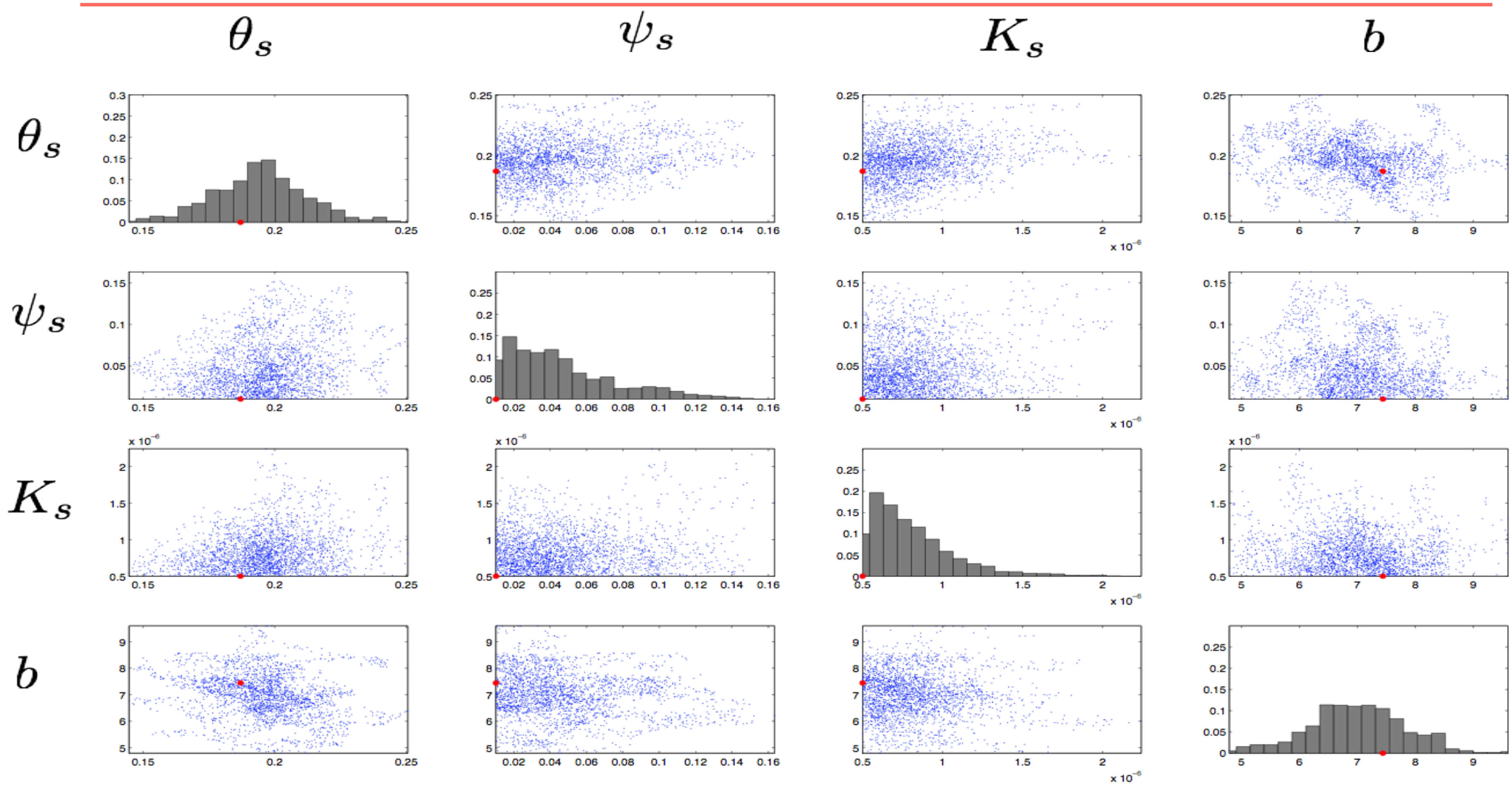
# Posterior parameter distributions $p(\theta|y)$




Samples in 2-parameter space show the reduction in uncertainty with the incorporation of observational information

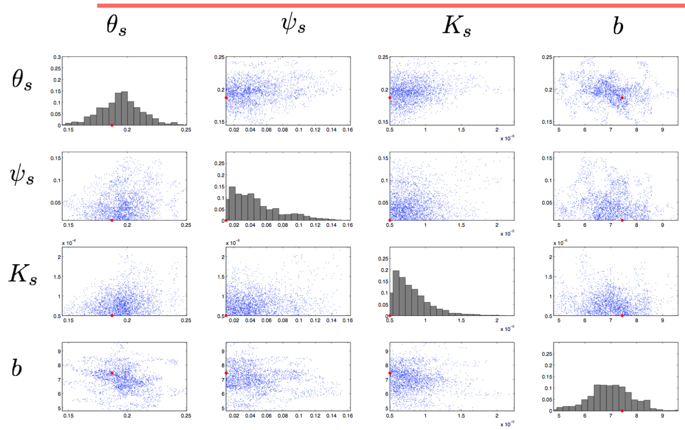
Red dots represent the PE solution

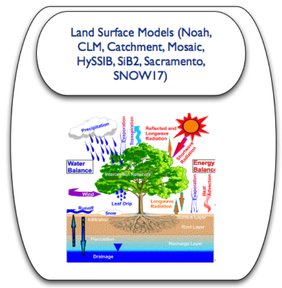
# Posterior parameter distributions $p(\theta|y)$




The parameter distribution reveal cross correlations between parameters, some stronger than others

# Probabilistic prediction

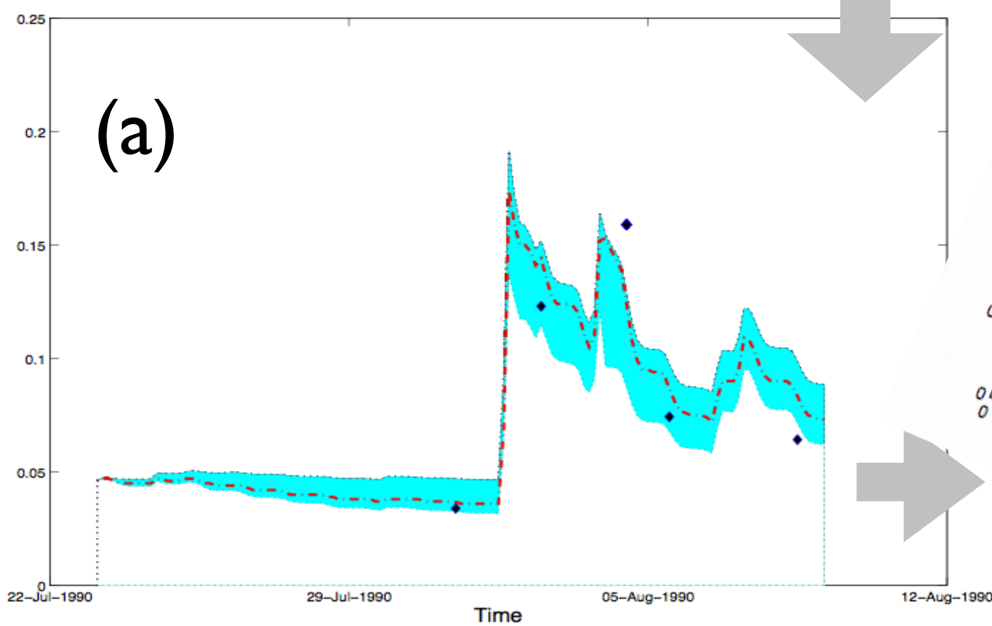


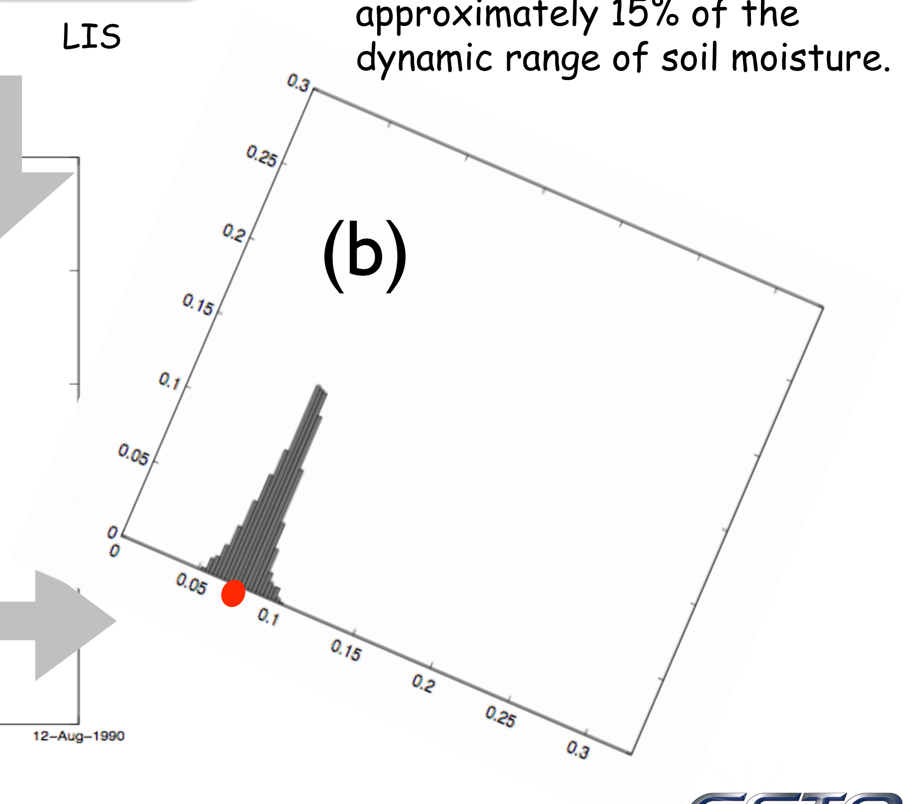




LIS

(a) The soil moisture time series with the PE solution (dashed line) and the 5th and 95th percentiles of the predictive distribution.

(b) Predictive distribution at the last time step; the spread is approximately 15% of the dynamic range of soil moisture.






Earth Science Technology Office

# Summary

- A flexible infrastructure for optimization and uncertainty modeling has been developed with the Land Information System.
- The infrastructure supports multiple optimization algorithms, different types of optimization problems and different types of objective function evaluation approaches/metrics.
- The implementation of the algorithms has been verified using test cases of varying complexity.
- The development of the optimization subsystem enables increased exploitation of the information content from observations.